Winning the TAC KBP Slot Filling Challenge

Ang Sun

Director of Research, Principal Scientist, inome
asun@inome.com

Outline

- The Slot Filling Challenge
- Overview of the NYU 2011 System
- Pattern Filler
- Distant Learning Filler

The Slot Filling Challenge

- Hand annotation performance
  - Precision: 70%
  - Recall: 54%
  - F-measure: 61%
- Top systems rarely exceed 30% F-measure

The Slot Filling Challenge

Query:
<query id="SF114">
<name>Jim Parsons</name>
<docid>eng-WL-11-174592-12943233</docid>
<enttype>PER</enttype>
<nodeid>E0300113</nodeid>
<ignore>per:date_of_birth, per:age, per:city_of_birth</ignore>
</query>

DOC1000001:
After graduating from high school, Jim Parsons received an undergraduate degree from the University of Houston. He was prolific during this time, appearing in 17 plays in 3 years.

Response:
SF114 per:schools_attended University of Houston

The Slot Filling Challenge

- Entry level is pretty high

  Jim Parsons was born and raised in Houston …
  … He attended Klein Oak High School in …

  - High performance name extraction
  - High performance coreference resolution
  - … …
- Extraction at large scale
  - 2011: 1.8 million documents
  - 2012: 3.7 million documents

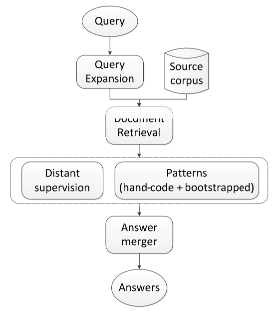
The Slot Filling Challenge

- Documents have not gone through a careful selection process
  - Evaluation in a real world scenario
- Slot types are of different granularities
  - per:employee_of
  - org: top_members/employees
  - … …

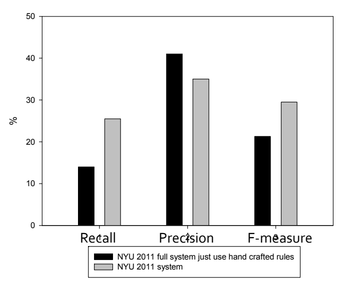
The Slot Filling Challenge

Person		Organization
per:alternate_names	per:title	org:alternate_names
per:date_of_birth	per:member_of	org:political/religious_affiliation
per:age	per:employee_of	org:top_members/employees
per:country_of_birth	per:religion	org:number_of_employees/members
per:stateorprovince_of_birth	per:spouse	org:members
per:city_of_birth	per:children	org:member_of
per:origin	per:parents	org:subsidiaries
per:date_of_death	per:siblings	org:parents
per:country_of_death	per:other_family	org:founded_by
per:stateorprovince_of_death	per:charges	org:founded
per:city_of_death		org:dissolved
per:cause_of_death		org:country_of_headquarters
per:countries_of_residence		org:stateorprovince_of_headquarters
per:stateorprovinces_of_residence		org:city_of_headquarters
per:cities_of_residence		org:shareholders
per:schools_attended		org:website

Overview of the NYU 2011 System



Pattern Filler



Pattern Filler

- Hand crafted patterns

pattern set	patterns	slots
local patterns for *person* queries	title of org, org title, org's title, title	title, employee_of
	title in GPE, GPE title	origin, location_of_residence
	person, integer,	age

Pattern Filler

- Hand crafted patterns

pattern set	patterns	slots
local patterns for *person* queries	title of org, org title, org's title, title	title, employee_of
	title in GPE, GPE title	origin, location_of_residence
	person, integer,	age
local patterns for org queries	title of org, org title, org's title	top_members/employees
	GPE's org, GPE-based org, org of GPE, org in GPE	location_of_headquarters
	org's org	subsidiaries / parent
implicit organzation	title [where there is a unique org mentioned in the current + prior sentence]	employee_of [for person queries]; top_members/employees [for org queries]
functional noun	F of X, X's F where F is a functional noun	family relations; org parents and subsidiaries

Pattern Filler

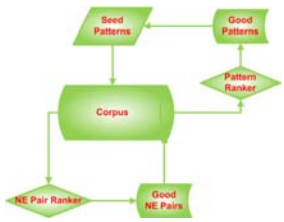
- Hand crafted patterns

```
// --- patterns with title and organization ---
//      (president of Ford, Ford President)

org-title-pattern  := org-title-pattern1 | org-title-pattern2;

org-title-pattern1 := (full-title):FullTitle "of" [constit cat=name pa=[head=ORGANIZATION]]:Org;

org-title-pattern2 := [constit cat=name pa=[head=ORGANIZATION]]:Org "'s"? (full-title):FullTitle;

full-title         := title-mod*
                        ([constit cat=n KBPtarget=true pa=[head?isa(titleOrOccupation)]] |
                         [constit cat=title KBPtarget=true pa=[head?isa(titleOrOccupation)]]);

title-mod          := [constit cat=n];

// note value (FullTitle) is a span!
when org-title-pattern
                   add [slot slotName=title value=FullTitle],
                   add [slot slotName=employee of value=Org];
```

- http://cs.nyu.edu/grishman/jet/jet.html

Pattern Filler

- Learned patterns (through bootstrapping)



Basic Idea:
It starts from some seed patterns which are used to extract named entity (NE) pairs , which in turn result in more semantic patterns learned from the corpus.

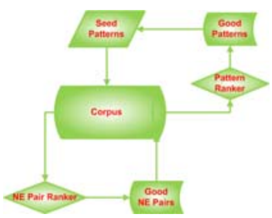
Pattern Filler

- Learned patterns (through bootstrapping)

“, chairman of ”

Pattern Filler

- Learned patterns (through bootstrapping)

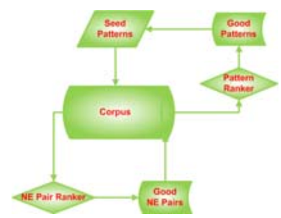


“, chairman of ”

<Bill Gates, Microsoft>, <Steve Jobs, Apple > …

Pattern Filler

- Learned patterns (through bootstrapping)

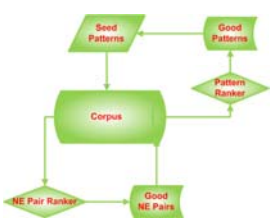


“, CEO of ”, “, director at”, … …

<Bill Gates, Microsoft>, <Steve Jobs, Apple > …

Pattern Filler

- Learned patterns (through bootstrapping)

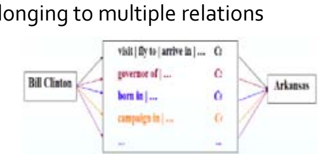


“, CEO of ”, “, director at”, … …

<Jeff Bezos, Amazon>, … …

Pattern Filler

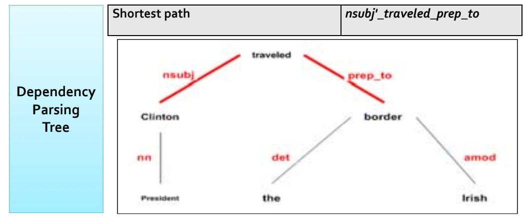
- Learned patterns (through bootstrapping)
  - Problem: semantic drift
  - a pair of names may be connected by patterns belonging to multiple relations



Pattern Filler

- Learned patterns (through bootstrapping)
  - Problem: semantic drift
  - Solutions:
    - Manually review top ranked patterns
    - Guide bootstrapping with pattern clusters

Pattern Filler



<e1>*President Clinton*</e1> traveled to <e2>*the Irish border*</e2> for an evening ceremony.

Distant Learning Filler

- **Distant Learning** (the general algorithm)
  - **Map** relations in knowledge bases to KBP slots
  - **Search** corpora for sentences that contain name pairs
  - **Generate** positive and negative training examples
  - **Train** classifiers using generated examples
  - **Fill** slots using trained classifiers

Distant Learning Filler

- **Distant Learning**
  - **Map** 4.1M Freebase relation instances to 28 slots
  - Given a pair of names ***<i,j>*** occurring together in a sentence in the KBP corpus, treat it as a
    - **positive** example if it is a Freebase relation instance
    - **negative** example if ***<i,j>*** is not a Freebase instance but ***<i,j'>*** is an instance for some **j'≠j**.
  - **Train** classifiers using MaxEnt
  - **Fill** slots using trained classifiers, in parallel with other components of NYU system

Distant Learning Filler

- **Problems**
  - **Problem 1**: Class labels are **noisy**
    - **Many False Positives** because name pairs are often connected by non-relational contexts

Freebase	Organization	Founded by
	Microsoft	Bill Gates

Example	Class Label
Bill Gates *has declared war on* ***Microsoft****'s insecure software*	FALSE POSITIVES
... ***Microsoft*** *,* ***Bill Gates****' relationship with India*	
Bill Gates *and* ***Microsoft*** *need to find some way to*	
... ...	

Distant Learning Filler

- **Problems**
  - **Problem 1**: Class labels are **noisy**
    - **Many False Negatives** because of incompleteness of current knowledge bases

Attribute of Person in Freebase	Incompleteness
place_of_birth	0.792
places_lived	0.923
nationality	0.786
parents	0.988
education	0.938
employment_history	0.966
... ...	... ...

$$\text{Incompleteness(Attr.)} = \frac{\text{\# Person without Attr.}}{\text{\#Person}}$$

Distant Learning Filler

- **Problems**
  - **Problem 2**: Class distribution is **extremely unbalanced**
    - Treat as negative if ***<i,j>*** is NOT a Freebase relation instance
      - Positive VS negative: **1:37**
    - Treat as negative if
      <i,j> is NOT a Freebase instance but ***<i,j'>*** is an instance for some **j'≠j AND *<i,j>*** is separated by no more than 12 tokens
      - Positive VS negative: **1:13**
    - Trained classifiers will have low recall, biased towards negative

Distant Learning Filler

- **Problems**
  - **Problem 3**: training ignores co-reference info
    - Training relies on full name match between Freebase and text
    - But partial names (*Bill*, *Mr. Gates* …) occur often in text
    - Use co-reference during training?
      - Co-reference module itself might be inaccurate and adds noise to training
    - But can it help during testing?

Distant Learning Filler

- **Solutions to Problems**
  - **Problem 1**: Class labels are **noisy**
    - Refine class labels to reduce noise
  - **Problem 2**: Class distribution is **extremely unbalanced**
    - Undersample the majority classes
  - **Problem 3**: training ignores co-reference info
    - Incorporate coreference during testing

Distant Learning Filler--Class Label Refinement

- The refinement algorithm
  I. Represent a training instance by its dependency pattern, the shortest path connecting the two names in the dependency tree representation of the sentence
  II. Estimate precision of the pattern

$$prec(p,c_i) = \frac{count(p,c_i)}{\sum_j count(p,c_j)}$$

  Precision of a pattern ***p*** for the class ***C_i*** is defined as
  the number of occurrences of ***p*** in the class ***C_i*** divided by
  the number of occurrences of ***p*** in any of the classes ***C_j***

  III. Assign the instance the class that its dependency pattern is most precise about

Distant Learning Filler--Class Label Refinement

- The refinement algorithm (cont)
  - Examples

Freebase	PERSON	Employee of
	Jon Corzine	Goldman Sachs

Freebase	Organization	Founded by
	CBS	William S. Paley

Example Sentence			Class
Jon Corzine	*appos chairman prep_of*	Goldman Sachs	PERSON: Employee_of
William S. Paley	*appos chairman prep_of*	CBS … …	ORG: Founded_by

prec(*appos chairman prep_of*, ***PERSON:Employee_of***) = 0.754
prec(*appos chairman prep_of*, ***ORG:Founded_by***) = 0.012

Distant Learning Filler--Undersampling the Majority Classes

- **Effort 1**:
  multiple n-way instead of ***single n-way*** classification
  - ***single n-way:*** an n-way classifier for all classes
    - Biased towards majority classes
  - ***multiple n-way***: an n-way classifier for each pair of name types
    - A classifier for PERSON and PERSON
    - Another one for PERSON and ORGANIZATION
    - … …
  - On average (10 runs on 2011 evaluation data)
    - ***single n-way:*** **180** fills for **8** slots
    - ***multiple n-way:*** **240** fills for **15** slots

Distant Learning Filler-- Undersampling the Majority Classes

- **Effort 2**:
  - Even with multiple n-way classification approach
  - *OTHER* (not a defined KBP slot) is still the majority class for each such n-way classifier
  - Downsize *OTHER* by randomly selecting a subset of them

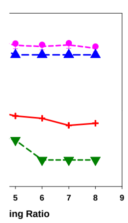
Distant Learning Filler-- Contribution of Coreference

- No use of co-reference during training
- Run Jet (NYU IE toolkit) to get co-referred names of the query
- Use these names when filling slots for the query
- Co-reference is beneficial to our official system
  - **P/R/F** of the distant filler itself
    - **With** co-reference: **36.4/11.4/17.4**
    - **Without** co-reference: **28.8/10.0/14.3**

Distant Learning Filler--Experimental Results (2011 evaluation data)

▪***Multiple n-way*** **outperformed** ***single n-way***

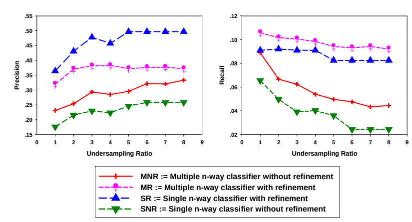
▪**Models with refinement:**
➢**higher performance**
➢**curves are much flatter**
➢**less sensitive to undersampling ratio**
➢**more robust to noise**



Undersampling Ratio: ratio between negatives and positives

MNR := Multiple n-way classifier without refinement
MR := Multiple n-way classifier with refinement
SR := Single n-way classifier with refinement
SNR := Single n-way classifier without refinement

Distant Learning Filler--Experimental Results (2011 evaluation data)



➢ **Models with refinement have better *P, R***
➢ ***Multiple n-way*** **outperforms** ***single n-way*** **mainly through improved recall**

Thanks!

?

Overview of 2011 System

- **Baseline: 2010 System** (three basic components)
  1) **Document Retrieval**
     - Use Lucene to retrieve a maximum of 300 documents
     - Query: the query name and some minor name variants
  2) **Answer Extraction**
     - Begins with text analysis: POS tagging, chunking, name tagging, time expression tagging, and coreference
     - Coreference is used to fill *alternate_names* slots
     - Other slots are filled using patterns (hand-coded and created semi-automatically using bootstrapping)
  3) **Merging**
     - Combines answers from different documents and passages, and from different answer extraction procedures

Overview of 2011 System

- **Passage Retrieval (QA)**
  - For each slot, a set of index terms is generated using distant supervision (using Freebase)
  - Terms are used to retrieve and rank passages for a specific slot
  - An answer is then selected based on name type and distance from the query name
  - Due to limitations of time, this procedure was only implemented for a few slots and was used as a fall-back strategy, if the other answer extraction components did not find any slot fill.

Overview of 2011 System

- Result Analysis (**NYU2** *R/P/F 25.5/35.0/29.5*)

module	score using only module			score excluding module		
	recall	precision	F1	recall	precision	F1
distant sup[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	35.4	25.7
distant sup[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	34.5	26.2
distant sup[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	35.7	29.2
alternate n[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	34.1	27.5
local patter[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	33.4	23.8
implicit organization	[illegible]	[illegible]	[illegible]	[illegible]	39.2	30.5
functional nouns	0.5	23.8	1.0	25.1	35.3	29.3
bootstrapped linear patterns	3.5	54.1	6.6	24.8	34.6	28.9
bootstrapped dependency patterns	1.8	36.2	3.4	25.0	35.2	29.2

	Recall	Precision	F1
NYU1 (With QA)	25.7	33.6	29.1
NYU2 (Without QA)	25.5	35.0	29.5

Performance of NYU Systems